



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

The Nature and Function of Educational Measurements

S. A. COURTIS, *Director of Instruction, Teacher Training and Research, Detroit, Mich., Public Schools*

(Continued from October issue)

Now, what about music? Are you all of one mind about how music should be taught? I fear not. One school tells me that the only way to achieve worth while results in music is to put a great deal of time on technique. Another school says the song approach is the best way. Still another school is equally insistent that appreciation is the "open sesame" to the development of musical talent. What is the non-expert to do? All three schools cannot be right. How are such differences to be harmonized? Is each one to try to put his own opinion over by brute force and the power of a "book company" organization, or should all submit their claims to the acid test of scientific investigation and modify their opinions on the basis of the actual facts? This point is so important that it needs direct illustration. The scientific method makes for progress because it enables open-minded men to get together on the basis of fact.

For instance, in this figure, what do you think you are looking at? Spirals? Nonsense! The lines are all perfect circles. Is my statement convincing? Do

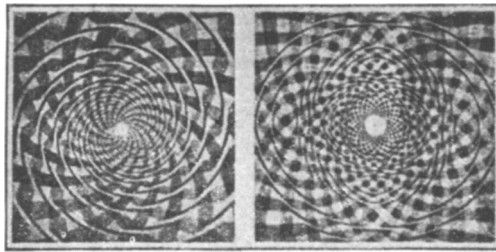


FIGURE 7

you find that just because I state my opinion emphatically, you see the lines less as spirals and more as circles? Ask yourself just how much oratory on my part and how many hours of debate it would take to change your opinion.

If, however, instead of wasting breath in oratory or in attacking the other fellow's opinion, we resort to measurement, we can soon reach perfect agreement. Watch me, while with this pointer, I follow around this line which seems to run into the center. You see, I come right back to the starting point. Better still, try it yourself. As the evidence piles up, don't you feel your opinion crumbling? Wouldn't the time come when you would be ready to say, "I can see that something is the matter? The results show that the lines really are circles, but I still continue to see them as spirals. How is that to be explained?" Now you are in quite a different frame of mind. You are open to the evidence which the investigations of the psychologists have given us that the background and these peculiar twisted-cord lines affect the nerves of the eye and produce the illusion of spirals. So it has proved in other subjects in which we have measured and so it will undoubtedly prove in music. Opinions will change in the light of facts. Truth is still mighty and will prevail. But, we must give truth a chance. The greater, the more fundamental, the differences of opinion among you, the greater the obligation resting upon all of you to submit your case to the court of final appeal,—open-minded, honest, scientific experimentation.

"But," you will say, "all the music tests I have seen measure such mechanical, such trivial details in music. They never get anywhere near the vital elements.

Why should I use the time I might be giving to really important matters to fuss with unessential mechanical details?" And I can see by the looks on your faces that many of you consider that a final and unanswerable argument.

Well, let's go into other fields and see how measurement operates there. For instance, how would you measure an elephant? What do you mean by the measurement of an elephant, anyway? Of course, if you were building a cage for an elephant, you would measure his height in order to know how high to make the cage, but you wouldn't consider you had measured the whole elephant, would you? Even when you had weighed the elephant, in order to tell how large timbers to put in the floor of the cage, or measured his strength in order to choose bars of sufficient size to hold him in, you would not have measured the elephant as a whole. Yet, the three partial measurements of the elephant you did make would prove very servicable in helping you to deal with the "unmeasured whole" intelligently. Now, note particularly that it is possible to measure the height, weight, and strength of a boy also, and that the elephant's measurements would be many times larger than the boy's. Would you, therefore, conclude that the elephant was more important than the boy, or would you be sensible and infer that when a creature so small can control one so much larger there must be something in the boy much more important than anything in the elephant?

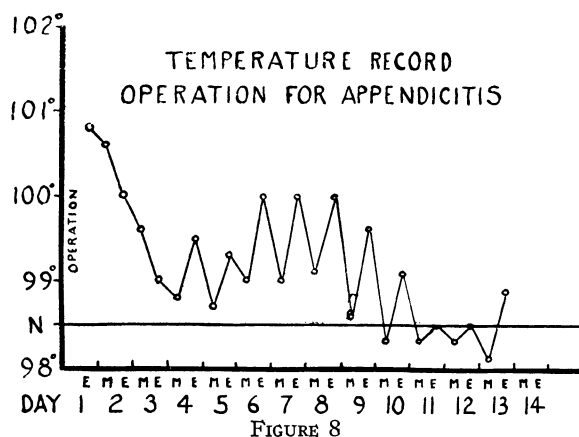
Do you see what I mean? Measurement is a tool, a means to an end. Tests and scores have no value in and of themselves. Their value is always to be sought in the use that is made of the results obtained by their use. The result of any measurement is never a complete evaluation of any situation. In every measurement that was ever made, there were other, larger, more "intangible" elements of importance present which for the time being and for the narrow purpose in view were disregarded. This, however, in no way vitiates the measurement or takes from its value: for it often happens that exact knowledge of a single and apparently trivial factor will serve as a significant guide for the control of the entire situation.

For instance, in Detroit one morning not so long ago, a lady awoke in violent pain. She had not been feeling just herself for several days, but there was no mistaking the fact this morning that something serious was the matter. She sent for her physician, and told him she believed her very life was in danger. Then a very strange thing happened; at least it would seem strange, if we were not accustomed to it, and did not know what it meant. The physician gave that lady a number of what to a non-intelligent onlooker would appear to be the most trivial and unrelated test. First he said, "Let me see your tongue." Next, he put a little glass rod in her mouth for a minute, then carefully measured the length of the mercury in the rod. Next, he took out his watch and gave her the kind of a timed test we call "taking the pulse rate." There were many other similar simple tests, yet when he had finished, his conclusion was "You have a clear case of acute appendicitis and I will not be responsible for the result unless you submit to an immediate operation." And in less than two hours the offending portion of the lady's anatomy had been removed.

Now in the modern hospital they don't take any more chances than necessary. They have learned that even trivial tests may prove an aid in controlling the great issues of life and death, and they keep careful records of their measurements. This is the record of the patient in whom we are interested.

On the evening of the first day her temperature was 101°. By the next morning it was lower and if you were that lady's surgeon you would watch this curve with greatest interest. Sometimes the curve goes down, down past the normal line and still down until the patient's life goes out. In this case, however, the reaction the fourth day is an indication that all is going well.

Note the variation the next few days. Why, the patient never made the

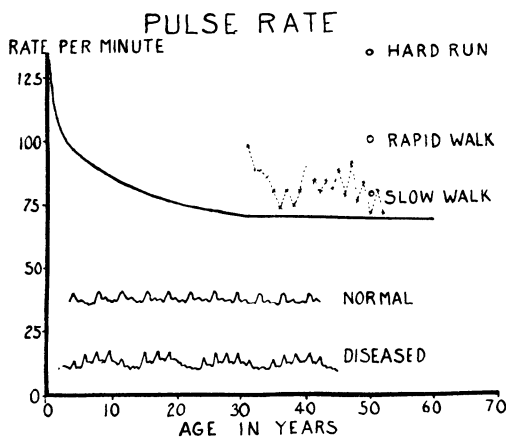


same score twice! Isn't that scandalous? The physician's thermometer must have been "no good"! At least that is the conclusion some teachers draw about our educational tests under similar conditions. Nevertheless, in due time, variation ceased, the curve dropped to normal level and the lady was sent home cured.

The moral is clear, isn't it? If a physician can use so insignificant a thing as the length of a tiny thread of mercury in a glass tube to guide him in his efforts to save life, may it not be just possible that some trivial test in music will be of similar aid to you in achieving the vital thing you call the spirit or soul of music. It is merely a question of discovery and the discovery will come in music, just as it has in other fields, when some one cares enough about his subject to pay the price in labor and experimentation.

I want to stress the fact that we measurement men have no mistaken ideas about the value of tests: the misunderstandings are all on the other side. We regard tests as tools created for a definite purpose, and serving that purpose admirably when properly used. We do not take results at their face value, nor do we let the test results master us. We keep them in subordination to our main purposes.

Here is the curve for pulse rate. A new born babe has a pulse rate of about 130. During the early years of life the rate falls off rapidly, until by 50, my age, the average is about 70 beats a minute. If I walk, my pulse will go up, say to 85. If I walk rapidly, it might go to 105. Suppose I had been running



violently for a long time and the very first thing upon my coming into this hall you took my pulse rate. If you found it was 130, would you infer, therefore, I was a new born babe? Yet sometimes teachers misuse the results of tests in ways which are just as foolish as that.

A test never reveals causes, it merely discloses conditions. In the lower part of Figure 11, there is the record of a normal pulse and just below it an abnormal record. You do not need to be a physician to know that something is the matter in the second case, but the record will not make the diagnosis for you. It is the function of tests to disclose the facts, but it is the function of the teacher to tell what the facts mean. If you were an experienced physician, the irregular record in the figure would mean a serious case of hardening of the liver. As a matter of fact, the patient died a day or two after this record was taken. But there is nothing in the record to tell the inexperienced person what the trouble was.

Consider another illustration. Here are charts showing the distribution of children about four schools. In each case the circles are drawn with the

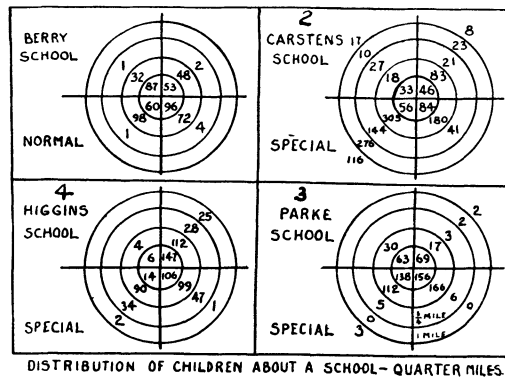


FIGURE 10

schools at the center. The circles represent distances of a quarter of a mile and the figures show the number of children living within each sector.

The Berry School represents a desirable condition. The children are evenly distributed about the school and very few live more than a half mile away. The other three schools show irregular distributions, all of the same type. Very few children are found in one sector.

Can you tell the explanation from these results? Most certainly not! The facts themselves are perfectly clear, but the interpretation to be put upon the facts can only be told by a further investigation.

In Figure 11, the results of such an investigation are given. In a map of the city of Detroit, a black headed pin has been placed for every five children. The Berry School is circle No. 1. See how evenly the children are distributed about the school. School No. 2 is the Cartens. Note that it is in a new district. There are open fields to the north. All the houses are to the south. In a few years this condition will be totally changed by the growth of the city. The Parke School is No. 3. The spaces to the north and east are the Packard and the Dodge automobile factories. Note that although the diagrams for these two schools are about the same, the conditions giving rise to the results are very different. So in the case of the third school. It is next door to a grave yard. The explanation differs from school to school. The statistical results point out with absolute certainty and clearness which schools vary from normal, but the value of such investigation depends, not upon these results but upon what is done about the condition revealed.

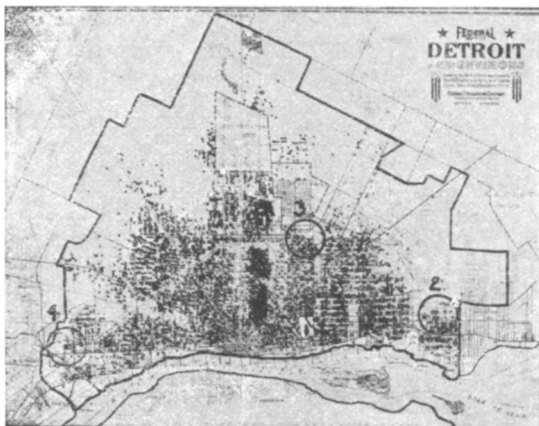


FIGURE 11

A second illustration will make the necessity for interpretation even more clear (Figure 12). One year we had an experiment in spelling in Detroit, in which about five hundred teachers took part. We were trying to determine whether or not with a very short course of study, we could teach every child to spell every word perfectly. There were only about sixty words to be learned in each grade. In the final test some classes did make 100%, while very many made 99%, 98%, etc. Yet out of the five hundred teachers who took part there were some who had class averages as low as 50%, and in these classes there would not be a single child who could spell every word perfectly.

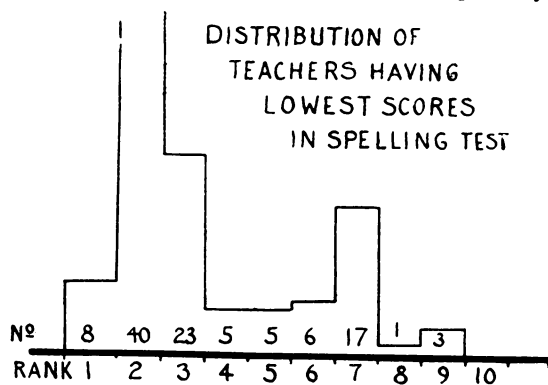


FIGURE 12

Do you suppose these classes were all taught by poor teachers? They certainly had very poor results. In Detroit we have an efficiency rating card, on which the principals report each year their opinion of the abilities of their teachers. We selected the one hundred twenty-five teachers whose classes made the lowest scores in the spelling tests and looked up their ratings (Figure 12). Eight were ranked as A No. 1 teachers, forty were in the next highest class and twenty-three in the third highest. There were a few teachers whose classes made low scores and whose ratings were also low, but that year if you had selected the seventy or eighty teachers whose classes made the lowest scores, you would have picked out the very finest teachers in Detroit.

The fact of the matter is that scores in a test do not tell anything about causes, and it is impossible to judge in terms of a single test whether a teacher is good or poor.

(To be continued)